

Long Arithmetic Progressions in Critical Sets

Ernie Croot ^{*}

February 1, 2008

Abstract

Given a density $0 < d \leq 1$, we show for all sufficiently large primes p that if $S \subseteq \mathbb{Z}/p\mathbb{Z}$ has the least number of three-term arithmetic progressions among all such sets having $\geq dp$ elements, then S must contain an arithmetic progression of length at least $\log^{1/4+o(1)} p$.

1 Introduction

Given a prime p , we say that $S \subseteq \mathbb{Z}/p\mathbb{Z}$ is a critical set for the density d if and only if $|S| \geq dp$ and S has the least number of three-term arithmetic progressions among all the subsets of $\mathbb{Z}/p\mathbb{Z}$ having at least dp elements. In this context, a three-term arithmetic progression is a triple of residue classes $n, n+m, n+2m$ modulo p . Note that this includes “trivial” progressions, which are ones where $m \equiv 0 \pmod{p}$, as well as “non-trivial” progressions, which are ones where $m \not\equiv 0 \pmod{p}$. We also distinguish two different progressions, according to how they are ordered: The progression $n, n+m, n+2m$ is considered different from $n+2m, n+m, n$.

The main result of the paper is the following theorem, which basically says that critical sets of positive density must have long arithmetic progressions.

Theorem 1 *Given $0 < d \leq 1$ we have that the following holds for all sufficiently large prime numbers p : If $S \subseteq \mathbb{Z}/p\mathbb{Z}$ is a critical set for the density d , then S must contain an arithmetic progression modulo p of length at least $\log^{1/4+o(1)} p$.*

^{*}Supported in part by an NSF grant.

Moreover, we show that for every $L \geq 1$, $0 < d \leq 1$, and p sufficiently large, there exists an arithmetic progression $N \subseteq \mathbb{Z}/p\mathbb{Z}$ having length at least $\log^L p$, such that

$$|S \cap N| > |N| \left(1 - \frac{1}{\log^{1/4+o(1)} p}\right).$$

We now compare this theorem with the state-of-the-art on long progressions in arbitrary sets of integers: As a consequence of W. T. Gowers's deep and beautiful proof of Szemerédi's Theorem [3, Theorem 18.6], one can show that for $0 < \delta \leq 1$, and all x sufficiently large, any set $S \subseteq \{1, 2, \dots, x\}$ having at least δx elements contains an arithmetic progression of length at least $\log \log \log \log(x) + c(\delta)$, for some constant $c(\delta)$. This is a considerably shorter AP than the one given for critical sets in our theorem above.

There are also some results for sumsets, which give much longer AP's. For example, J. Bourgain [1] proved the interesting result that if $A, B \subseteq \{1, \dots, x\}$, where $|A| > \delta x$, $|B| > \gamma x$, then $A + B$ has an arithmetic progression of length at least $\exp(c(\delta\gamma \log x)^{1/3} - \log \log x)$ (for some $c > 0$). I. Ruzsa [9] gave an ingenious construction, which shows that for every $0 < \epsilon < 1/3$, and all x sufficiently large, there exists a set A having at least $b(\epsilon)x$ elements (for some function $b(\epsilon) > 0$ that depends only on ϵ), such that $A + A$ has no arithmetic progressions longer than $\exp(\log^{2/3-\epsilon} x)$. Then, B. Green [4] improved Bourgain's result, and showed that a sumset $A + B$ has an arithmetic progression of length at least $\exp(c'(\delta\gamma \log x)^{1/2} - \log \log x)$. We note that the length of the progressions in these sumsets is much longer than the ones we give for critical sets; and so, if we could somehow prove that critical sets are sumsets of two large sets A and B , then our result could possibly be improved.

There are also some impressive results on long arithmetic progressions in repeated sumsets $A + A + \dots + A$ and subset sums, notably those of Freiman [2]; Sárkőzy [10], [11], and [12]; Lev [6], and [7]; Vu and Szemerédi [14] and [15]; and J. Solymosi [13].

Comments: The method of proof of our theorem has many common features with the result of B. Green [4]. In particular, we both make use of large deviation (or concentration of measure) results from probability theory; and we both use techniques involving Bohr neighborhoods. However, the combinatorial aspects of our two theorems are different, which reflects the fact that sumsets and critical sets have different properties that must be exploited in different ways.

It is possible to refine the proof of our theorem, to show that critical sets have AP's whose length depends on the density d ; so, for example, it might be possible to prove that critical sets $S \subseteq \mathbb{Z}/p\mathbb{Z}$ of density d have a long AP for any $d > (\log \log p)^{-1}$; and, if one applies Chang's Structure Theorem, as Green does, one can maybe get an even better result (longer AP's holding for lower densities d).

2 Proof of Theorem 1

We identify S with the indicator function $S(n)$, which is defined as follows:

$$S(n) = \begin{cases} 1, & \text{if } n \in S; \\ 0, & \text{otherwise.} \end{cases}$$

Next, we define the discrete Fourier transform of $S(n)$ to be

$$\hat{S}(a) = \sum_{0 \leq n \leq p-1} S(n) e^{2\pi i an/p}.$$

Then, we have that the number of 3-term arithmetic progressions in the set S is given by

$$\sum_{r+s \equiv 2t \pmod{p}} S(r)S(s)S(t) = \frac{1}{p} \sum_{0 \leq a \leq p-1} \hat{S}(a)^2 \hat{S}(-2a).$$

We now write this last sum as $\Sigma_1 + \Sigma_2$, where Σ_1 is the sum over all those a where

$$|\hat{S}(-2a)| > \frac{p \log \log p}{\sqrt{\log p}}, \quad (1)$$

and where Σ_2 is the sum over the remaining values of a . From Parseval's identity we deduce the estimate

$$|\Sigma_2| \leq \frac{p \log \log p}{\sqrt{\log p}} \sum_{(*)} |S(a)|^2 \leq \frac{dp^3 \log \log p}{\sqrt{\log p}}, \quad (2)$$

where the condition $(*)$ is that we sum over all $0 \leq a \leq p-1$ that do not satisfy (1).

We now bound the number of terms in Σ_1 from above: Denote this number of terms by M . Then, by Parseval's identity we get that

$$\frac{p^2 (\log \log p)^2}{\log p} M < \sum_{0 \leq a \leq p-1} |\hat{S}(a)|^2 = dp^2,$$

which implies

$$M < \frac{d \log p}{(\log \log p)^2}. \quad (3)$$

We next require the following basic lemma:

Lemma 1 *Suppose that $K \geq 1$, $0 \leq a_1, \dots, a_k \leq p-1$ and*

$$k < \frac{\log p}{2K \log \log p}.$$

Then, for p sufficiently large there is at least one integer $1 \leq n \leq p-1$ lying in the Bohr neighborhood defined by

$$\text{For all } i = 1, 2, \dots, k, \quad \left\| \frac{a_i n}{p} \right\| < \frac{1}{\log^K p}, \quad (4)$$

where $\|x\|$ denote the distance from x to the nearest integer.

Proof of the Lemma. This is nothing more than Dirichlet's pigeonhole argument: We consider the p vectors lying in the unit k -cube

$$(a_1 y/p \pmod{1}, \dots, a_k y/p \pmod{1}),$$

where y runs through the integers $0, 1, \dots, p-1$. Now, by the pigeonhole principle, there must exist two values of y , say y_1 and y_2 , such that $\|a_i(y_1 - y_2)/p\| < 1/\log^K p$. \blacksquare

Let $\{a_1, \dots, a_k\}$ be the values of a satisfying (1), which are the indices of the terms in Σ_1 . Then, we apply Lemma 1 with $K = 2L$, and deduce that there is an integer n_0 satisfying (4). Now, let N be the arithmetic progression

$$N = \{j n_0 \pmod{p} : 0 \leq j < \log^L p\}.$$

We identify N with its scaled indicator function

$$N(n) = \begin{cases} \frac{1}{|N|}, & \text{if } n \in N; \text{ and} \\ 0, & \text{if } n \notin N. \end{cases}$$

Then, we define the Fourier transform of this scaled indicator function:

$$\hat{N}(a) = \sum_{n=0}^{p-1} N(n) e^{2\pi i a n / p}.$$

We now consider the convolution

$$(S * N)(m) = \sum_{a+b \equiv m \pmod{p}} S(a)N(b) = \frac{1}{|N|} \sum_{n \in N} S(m-n).$$

It is obvious that

$$0 \leq (S * N)(m) \leq 1.$$

And, we have the following basic fact

Lemma 2 *Suppose that*

$$1 > \epsilon > \frac{1}{\log^L p}.$$

*If $(S * N)(m) > 1 - \epsilon$, for some $0 \leq m \leq p-1$, then S contains an arithmetic progression of length at least ϵ^{-1} .*

Proof of the Lemma. If $(S * N)(m) > 1 - \epsilon$, then we are saying that the set S contains all but $\epsilon \log^L p$ of the residues

$$m, m - n_0, m - 2n_0, \dots, m - \lfloor \log^L p \rfloor n_0 \pmod{p}. \quad (5)$$

Clearly, then, S will contain an AP of length at least ϵ^{-1} for $\epsilon > 1/\log^L p$.

■

We will now show that if S is a critical set, then

$$(S * N)(m) > 1 - \frac{\log \log p}{\log^{1/4} p}, \quad (6)$$

for some m ; and so, our theorem will follow from Lemma 2.

To show that this is the case, suppose, for proof by contradiction, that (6) fails to hold for every $0 \leq m \leq p-1$; and, let

$$\kappa = \max \left(1 - \frac{\log \log p}{\log^{1/4} p}, \max_{0 \leq m \leq p-1} |(S * N)(m)| \right).$$

Then, define the weighting function $w(m)$ for $0 \leq m \leq p-1$ to be

$$w(m) = \kappa^{-1} (S * N)(m).$$

Clearly,

$$0 \leq w(m) \leq 1;$$

Now we need the following lemma:

Lemma 3 Suppose that $w(m)$ is a real-valued function supported on the integers in $[0, p-1]$, satisfying $0 \leq w(m) \leq 1$. Then, there exists a function $u(m)$, also supported on the integers in $[0, p-1]$, such that

1. $u(m) \in \{0, 1\}$ for all $m = 0, 1, \dots, p-1$;
2. $\hat{u}(a) = \hat{w}(a) + O((\log p)\sqrt{p})$; and,
3. $\hat{u}(0) = \hat{w}(0) + \delta$, where $0 \leq \delta < 1$.

Before we can prove this lemma we require the following concentration of measure result due to Hoeffding [5] (also see [8], Theorem 5.7):

Proposition 1 Suppose that v_1, \dots, v_r is a sequence of independent random variables where $|v_i| < 1$. Let

$$\nu = E(v_1 + \dots + v_r) = E(v_1) + \dots + E(v_r),$$

and let $\Sigma = v_1 + \dots + v_r$. Then,

$$P(|\Sigma - \nu| > rt) \leq 4 \exp(-rt^2/2).$$

Remark: A stronger result is possible here, using Hoeffding's theorem. The result here is obtained as follows: Write $v_i = x_i + iy_i$, where $-1 \leq x_i, y_i \leq 1$, and then observe that the if the “bad event” $|\Sigma - \nu| > rt$ occurs, then either we have the “bad event” $|\Sigma_x - \nu_x| > rt/\sqrt{2}$ or the “bad event” $|\Sigma_y - \nu_y| > rt/\sqrt{2}$, where $\Sigma_x = x_1 + \dots + x_r$ and $\nu_x = E(x_1 + \dots + x_r)$, and where Σ_y and ν_y are defined analogously. Using Hoeffding's theorem, the probability that either of these last two bad events occurring is at most $4 \exp(-rt^2/2)$, as in the proposition above.

Proof of the Lemma. We will let $u'(m)$ be a sequence of independent Bernoulli random variables with distribution

$$P(u'(m) = 1) = w(m).$$

We note that

$$E(u'(m)) = w(m). \quad (7)$$

Then, for each integer a satisfying $0 \leq a \leq p-1$, we have that the Fourier transform

$$\hat{u}'(a) = \sum_{j=0}^{p-1} u'(j) e^{2\pi i j a / p}$$

can be interpreted as a sum of independent random variables as follows

$$\hat{u}'(a) = v_0 + \dots + v_{p-1}, \quad \text{where } v_j = u'(j) e^{2\pi i j a / p}.$$

Now,

$$E(\hat{u}'(a)) = E(v_0) + \cdots + E(v_{p-1}) = \sum_{j=0}^{p-1} E(u'(j))e^{2\pi i ja/p} = \hat{w}(a).$$

Applying the Hoeffding proposition above, we deduce that

$$P(|\hat{u}'(a) - \hat{w}(a)| \geq (\log p)\sqrt{p}) < 4 \exp(-(\log^2 p)/2).$$

Thus, the probability that

$$\text{For all } a = 0, 1, \dots, p-1, \quad |\hat{u}'(a) - \hat{w}(a)| < (\log p)\sqrt{p} \quad (8)$$

is at least

$$1 - 4p \exp(-(\log^2 p)/2),$$

which is positive for $p \geq 11$.

Since (8) holds with positive probability, there must exist a function $u(m)$, supported on $0, 1, \dots, p-1$, taking the values 0 and 1, and such that

$$\text{For all } a = 0, 1, \dots, p-1, \quad |\hat{u}(a) - \hat{w}(a)| < (\log p)\sqrt{p}.$$

Then, by reassigning at most $O((\log p)\sqrt{p})$ of the $u(m)$'s to 0 or 1 as needed, we can get

$$\hat{u}(0) = \hat{w}(0) + \delta, \quad 0 \leq \delta < 1,$$

while maintaining

$$\hat{u}(a) = \hat{w}(a) + O((\log p)\sqrt{p})$$

for all the other values $a = 1, 2, \dots, p-1$. Thus, we have constructed a function $u(m)$ which satisfies the conclusion of our lemma. ■

Now let S' denote the set for which $u(m)$ is the indicator function. Then, we have that

$$|S'| = \kappa^{-1}|S| + \delta, \quad \text{where } 0 \leq \delta < 1.$$

We now estimate the number of 3AP's contained in S' modulo p : This number is

$$\begin{aligned} \frac{1}{p} \sum_{a=0}^{p-1} \hat{u}(a)^2 \hat{u}(-2a) &= \frac{1}{p} \sum_{a=0}^{p-1} (\hat{w}(a) + O((\log p)\sqrt{p}))^2 (\hat{w}(-2a) + O((\log p)\sqrt{p})) \\ &= \frac{1}{p} \sum_{a=0}^{p-1} \hat{w}(a)^2 \hat{w}(-2a) + E, \end{aligned} \quad (9)$$

where

$$E = O \left(\frac{\log p}{\sqrt{p}} \sum_{a=0}^{p-1} (\log^2 p)p + (\log p)\sqrt{p}|w(a)| + |\hat{w}(a)|^2 + |\hat{w}(a)\hat{w}(-2a)| \right).$$

Using the Cauchy-Schwarz inequality, in combination with Parseval's identity, one can show that

$$E = O((\log^3 p)p\sqrt{p});$$

and so it follows that the number of 3AP's in S' modulo p is

$$\begin{aligned} \frac{1}{p} \sum_{a=0}^{p-1} \hat{w}(a)^2 \hat{w}(-2a) &+ O((\log^3 p)p\sqrt{p}) \\ &= \frac{1}{\kappa^3 p} \sum_{a=0}^{p-1} \hat{S}(a)^2 \hat{S}(-2a) \hat{N}^2(a) \hat{N}(-2a) + O((\log^3 p)p\sqrt{p}). \end{aligned} \tag{10}$$

We now break this last sum into the two sums $\Sigma'_1 + \Sigma'_2$, where Σ'_1 is over those $0 \leq a \leq p-1$ satisfying (1), and Σ'_2 is the sum for the remaining values of a . Now, for each a satisfying (1) and for each $n \in N$ we have from (4) with $K = 2L$ that

$$\left\| \frac{-2an}{p} \right\| \leq 2 \left\| \frac{an}{p} \right\| < \frac{2}{\log^L p}.$$

for p sufficiently large. The same estimate holds for the distance from an/p to the nearest integer. Thus,

$$\begin{aligned} \hat{N}(-2a) &= \frac{1}{|N|} \sum_{n \in N} e^{2\pi i (-2an)/p} \\ &= \frac{1}{|N|} \sum_{n \in N} \left(1 + O\left(\frac{1}{\log^L p}\right) \right) \\ &= 1 + O\left(\frac{1}{\log^L p}\right); \end{aligned}$$

and, the same estimate holds for $\hat{N}(a)$. Thus, we conclude that

$$\Sigma'_1 = \Sigma_1 + O\left(\frac{p^3}{\log^L p}\right).$$

We also have the estimate

$$|\Sigma'_2| \leq \frac{p \log \log p}{\sqrt{\log p}} \sum_{(*)} |\hat{S}(a)|^2 \leq \frac{dp^3 \log \log p}{\sqrt{\log p}},$$

where $(*)$ represents the condition that $0 \leq a \leq p-1$ such that a does not satisfy (1). We note that the inequality here follows from Parseval's identity.

Combining our estimate for Σ'_1 and Σ'_2 together with (10), we deduce that the number of 3AP's in S' modulo p is

$$\frac{1}{\kappa^3 p} (\Sigma'_1 + \Sigma'_2) + O((\log^3 p) p \sqrt{p}) = \frac{1}{\kappa^3 p} (\Sigma_1 + \Sigma_2) + O\left(\frac{dp^2 \log \log p}{\kappa^3 \sqrt{\log p}}\right).$$

Thus,

$$\begin{aligned} \#(\text{3AP}'s \text{ in } S' \pmod{p}) &= \frac{1}{\kappa^3} \times \#(\text{3AP}'s \text{ in } S \pmod{p}) \\ &\quad + O\left(\frac{dp^2 \log \log p}{\kappa^3 \sqrt{\log p}}\right). \end{aligned} \quad (11)$$

We now proceed to show that this is impossible, and from our chain of reasoning above, this would mean that (6) holds, and therefore the theorem would follow from Lemma 2.

To show the above equation cannot hold, we require the following combinatorial lemma, which is proved using the probabilistic method, in combination with the second moment method:

Lemma 4 *Suppose $A, B \subset \mathbb{Z}/p\mathbb{Z}$ have densities γ and δ , respectively; and, suppose that A and B contain $\alpha\gamma^3 p^2$ and $\beta\delta^3 p^2$ non-trivial 3AP's, respectively. Then, there exists a subset C of $\mathbb{Z}/p\mathbb{Z}$ having density at least*

$$\gamma\delta + O(p^{-1/4}),$$

such that the number of non-trivial 3AP's lying in C modulo p is at most

$$\alpha\beta(\gamma\delta)^3 p^2 + O(p^{3/2}).$$

Remark. The same result holds if we add in trivial AP's, since a subset D of $\mathbb{Z}/p\mathbb{Z}$ can have only $O(p)$ 3AP's, which is well within the error $O(p^{3/2})$.

Proof of Lemma 4. We will find a pair of integers u, v such that $A \cap (uB + v)$ has the desired properties. First, we show that this intersection has density very close to $\gamma\delta$ for almost all $0 \leq u, v \leq p-1$, by using a

second moment argument: We suppose that u and v are random variables chosen independently from $\{0, \dots, p-1\}$ with the uniform measure. Then, the variance $V(|A \cap (uB + v)|)$ is

$$E(|A \cap (uB + v)|^2) - E(|A \cap (uB + v)|)^2.$$

To compute the first expectation we express the intersection as a sum of indicator functions:

$$|A \cap (uB + v)| = \sum_{b \in B} f(ub + v),$$

where f is the indicator function for the set A . So, we have that

$$\begin{aligned} E(|A \cap (uB + v)|^2) &= \sum_{b, b' \in B} E(f(ub + v)f(ub' + v)) \\ &= \frac{1}{p^2} \sum_{b, b' \in B} \sum_{0 \leq u, v \leq p-1} f(ub + v)f(ub' + v). \end{aligned}$$

Now, given a pair of unequal elements $b, b' \in B$, and any two elements $a, a' \in A$ (a may equal a'), there is exactly one pair of numbers $u, v \pmod{p}$ which make $ub + v \equiv a \pmod{p}$ and $ub' + v \equiv a' \pmod{p}$. That is, we have that if $b' \neq b$, then there are exactly $|A|^2$ pairs u, v which make $f(ub + v)f(ub' + v) \neq 0$ (and therefore equal to 1). Thus,

$$E(|A \cap (uB + v)|^2) \leq \gamma^2 \delta^2 p^2 + |B|.$$

The term $|B|$ comes from those pairs b, b' with $b = b'$.

To estimate $E(|A \cap (uB + v)|)$, we note that for a fixed $b \in B$ and $0 \leq u \leq p-1$, the probability that $ub + v$ lies in A is γ . Thus, the expected size of this intersection is $\gamma \delta p$.

We now conclude that

$$V(|A \cap (uB + v)|) \leq |B| = \delta p;$$

and so, by an application of Chebychev's inequality we conclude that

$$P(|A \cap (uB + V)| < (1 - \epsilon)\gamma \delta p) \leq \frac{1}{\epsilon^2 \gamma^2 \delta p}.$$

Next, we compute the expected number of 3AP's in the intersection $A \cap (uB + v)$: Let $Q = Q(u, v)$ be the number of non-trivial 3AP's lying in $A \cap (uB + v)$. Now, suppose that x_1, x_2, x_3 is a non-trivial 3AP in A , so

that $x_2 \equiv x_1 + d, x_3 \equiv x_1 + 2d \pmod{p}$, for some $d \not\equiv 0 \pmod{p}$; and, suppose that y_1, y_2, y_3 is a non-trivial 3AP in B . Then, there is exactly one pair $0 \leq u, v \leq p-1$ such that

$$\text{For } i = 1, 2, 3, \quad ux_i + v \equiv y_i \pmod{p}.$$

Thus, for u, v chosen at random from $0, \dots, p-1$ with uniform probability, the probability that a particular non-trivial 3AP lies in $uB+v$ is $\beta\delta^3$; and so, the expected size of Q is $\alpha\beta(\gamma\delta)^3 p^2$. So, there can be at most $p^2 - p^{3/2}$ of the choices for u and v such that the intersection has more than $\alpha\beta(\gamma\delta)^3(p^2 + 2p^{3/2})$ 3AP's; else, if all but $p^{3/2}$ of the choices give more than this many 3AP's in this intersection, then we would have that Q exceeds

$$\frac{(p^2 - p^{3/2})(p^2 + 2p^{3/2})}{p^2} \alpha\beta(\gamma\delta)^3,$$

which we know is not the case. Thus, the probability that $Q < \alpha\beta(\gamma\delta)^3(p^2 + 2p^{3/2})$ is $> p^{-1/2}$. So, for $\epsilon = p^{-1/4}\gamma^{-1}\delta^{-1/2}$, we get that the events

$$|A \cap (uB + v)| \geq (1 - \epsilon)\gamma\delta p \quad \text{and} \quad Q < \alpha\beta(\gamma\delta)^3(p^2 + 2p^{3/2})$$

occur with positive probability. So, there is a choice for u and v so that both these events occur, which proves the lemma. ■

We require one more lemma before we can prove that (11) is impossible:

Lemma 5 *Given $0 < \theta < 1$, there exists a subset $U \subset \mathbb{Z}/p\mathbb{Z}$ having density $1 - \theta + O(1/p)$ such that the number of 3AP's lying in U , both trivial and non-trivial, is at most*

$$p^2(1 - 3\theta + 2.5\theta^2).$$

For $0 < \theta < 1/3$ this quantity is at most

$$p^2(1 - \theta)^3(1 - \theta^2/2).$$

Proof. First, we claim that the sum of the number of 3AP's (trivial and non-trivial) lying in U and lying in $\overline{U} = (\mathbb{Z}/p\mathbb{Z}) \setminus U$ is

$$p^2(1 - 3\theta + 3\theta^2). \tag{12}$$

This follows by inclusion-exclusion: The number of 3AP's lying in U is $x_1 - x_2 + x_3 - x_4$, where x_1 is the total number of 3AP's among the residue classes modulo p^2 ; x_2 is the sum of the number of these 3AP's $x, x+d, x+2d$ such that $x \in \overline{U}$, summed with the number where $x+d \in \overline{U}$, and summed

with the number where $x + 2d \in \overline{U}$; x_3 is the sum of the number of such progressions with $x, x+d \in \overline{U}$, then $x, x+2d \in \overline{U}$, and finally summed with the number where $x+d, x+2d \in \overline{U}$; finally, x_4 is the number of progressions in \overline{U} . It is easy to see that $x_1 = p^2$, $x_2 = 3rp^2$, and $x_3 = 3r^2p^2$. So, sum of the number of 3AP's in U and \overline{U} equals the expression in (12).

Now consider the set \overline{U} , having density $\theta + O(1/p)$, given as follows:

$$\overline{U} := [0, \theta p/2] \cup [p/2, p/2 + \theta p/2],$$

where here we take the integers in these two intervals (since U and \overline{U} are sets of residue classes modulo p). Call the first interval I_0 , and the second I_1 . If $x, y \in I_i$, then $z \equiv 2^{-1}(x+y) \pmod{p}$ lies in I_i if x, y have the same parity, and lies in I_{1-i} if they are of different parity. This gives that the number of 3AP's $x, x+d, x+2d$ is at least the number of ordered pairs x, y where both $x, y \in I_0$ or both are in I_1 . So, the number of 3AP's in \overline{U} is at least $\theta^2 p^2/2$, and it follows that the number of 3AP's in U is at most

$$p^2(1 - 3\theta + 3\theta^2 - \theta^2/2) = p^2(1 - 3\theta + 2.5\theta^2),$$

which proves the lemma. ■

Now we let $\theta = 1 - \kappa$, and let U be the set given by this lemma. Then, we apply Lemma 4 with $A = U$, and $B = S'$, and we deduce that there is a set C with

$$|C| = |S| + O(p^{3/4}),$$

such that C contains at most

$$\begin{aligned} \kappa^3 \left(1 - \frac{(1-\kappa)^2}{2}\right) \left(\frac{\#(\text{3AP's in } S)}{\kappa^3} + O\left(\frac{\log \log p}{\sqrt{\log p}}\right)\right) \\ = \#(\text{3AP's in } S) \left(1 - \frac{(\log \log p)^2}{2\sqrt{\log p}} + O\left(\frac{\log \log p}{\sqrt{\log p}}\right)\right). \end{aligned}$$

To show that this is impossible for sufficiently large p , we let C' be any set gotten from C by adding or removing at most $O(p^{3/4})$ elements such that

$$|C'| = |S|.$$

Then, in the worst case, each element we add to C (to produce C') adds at most p new 3AP's. Thus,

$$\begin{aligned} \#(\text{3AP's in } C') &= \#(\text{3AP's in } C) - O(p^{1.75}) \\ &< \#(\text{3AP's in } S) \left(1 - \frac{(\log \log p)^2}{2\sqrt{\log p}} + O\left(\frac{\log \log p}{\sqrt{\log p}}\right)\right). \end{aligned} \tag{13}$$

To get this inequality we have used a corollary of the following theorem of Varnavides [16], which allows us to absorb the error term $O(p^{1.75})$ into the error $O((\log \log p)/\sqrt{\log p})$:

Theorem 2 *Given $0 < \alpha \leq 1$, there exists $0 < c \leq 1$ such that for any set $T \subseteq \{1, 2, \dots, x\}$ having $|T| \geq \alpha x$,*

$$\#\{a, b, c \in T : a + b = 2c\} > cx^2.$$

This corollary is:

Corollary 1 *There exists $0 < c \leq 1$, depending only on d (the lower bound for the density of S), such that*

$$\#\{\text{3AP's in } S\} > cp^2.$$

The proof of this corollary is immediate, since if we think of S as a set of integers, say $S \subseteq \{0, 1, \dots, p-1\}$ (instead of as a set of residue classes modulo p), then every solution to $a + b = 2c$, $a, b, c \in S$ in the integers gives a solution $a + b \equiv 2c \pmod{p}$. So, the number of 3AP's in S modulo p is at least the number of 3AP's in S , when we think of it as a subset of the integers.

Now, (13) contradicts the fact that S is a critical set: Here we have constructed a set C' having the same cardinality as the set S , but where C' has fewer 3AP's than S . Thus, we must conclude that

$$c(n) > 1 - \frac{\log \log p}{\log^{1/4} p}$$

for some $0 \leq n \leq p-1$, and the theorem is proved. \blacksquare

References

- [1] J. Bourgain, *On Arithmetic Progressions in Sums of Sets of Integers*, A Tribute to Paul Erdős, 105-109, Cambridge University Press, Cambridge, 1990.
- [2] G. A. Freiman, H. Halberstam, and I. Ruzsa, *Integer Sums sets Containing Long Arithmetic Progressions*, J. London Math. Soc. (2) **46** (1992), 193-201.

- [3] W. T. Gowers, *A New Proof of Szemerédi's Theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465-588.
- [4] B. Green, *Arithmetic Progressions in Sumsets*, Geom. Funct. Anal. **12** (2002), 584-597.
- [5] W. Hoeffding, *Probability Inequalities for Sums of Independent Random Variables*, J. Amer. Statist. Assoc. **58** (1963), 13-30.
- [6] V. F. Lev, *Optimal Representations by Sumsets and Subset Sums*, J. Number Theory **62** (1997), 127-143.
- [7] ———, *Blocks and Progressions in Subset Sums Sets*, Acta Arith. **106** (2003), 123-142.
- [8] C. McDiarmid, *On the Method of Bounded Differences*, London Math. Soc. Lecture Note Ser. 14, Cambridge Univ. press, Cambridge, 1989.
- [9] I. Ruzsa, *Arithmetic Progressions in Sumsets*, Acta Arith. **60** (1991), no. 2, 191-202.
- [10] A. Sárkőzy, *Finite Addition Theorems. I*, J. Number Theory **32** (1989), 114-130.
- [11] ———, *Finite Addition Theorems. II*, J. Number Theory **48** (1994), 197-218.
- [12] ———, *Finite Addition Theorems. III*, Groupe de Travail en Théorie Analytique et Élémentaire des Nombres, 1989-1990, 105-122, Publ. Math. Orsay, 92-01, Univ. Paris XI, Orsay, 1992.
- [13] J. Solymosi, *Arithmetic Progressions in Sets with Small Sumsets*, Manuscript.
- [14] E. Szemerédi and V. Vu, *Long Arithmetic Progressions in Sum-Sets and the Number of x -sum-free Sets*, submitted.
- [15] ———, *Finite and Infinite Arithmetic Progressions in Sumsets*, submitted.
- [16] P. Varnavides, *On Certain Sets of Positive Density*, J. London Math. Soc. **34** (1959), 358-360.